**CSCI3656: NUMERICAL COMPUTATION**
**Homework 9: Due Friday, Nov. 5, 5:00pm**

Turn in your own writeup that includes your code. List any resources you used including collaborating with others. Submit a PDF on Canvas by Friday, Nov. 5 at 5pm.

1. Consider the function
$$f(x) = \sin(2\pi x) + \cos(3\pi x), \quad x \in [-1, 1]. \tag{1}$$
Compute the coefficients of a least-squares-fit degree-7 polynomial from $n = 33$ evenly spaced points. In other words, your training data are pairs $(x_i, y_i)$ with $i = 1, \ldots, 33$ where the $x_i$'s are evenly spaced points in $[-1, 1]$ (like, `linspace`) and $y_i = f(x_i)$. Make a plot of both $f(x)$ and the degree-7 polynomial approximation.

2. Create *testing data* by (i) choosing 100 random points ($x_i'$, $i = 1, \ldots, 100$) in the interval $[-1, 1]$ and (ii) evaluating $f(x)$ at each of those points ($y_i' = f(x_i')$). This gives you a new set of data $(x_1', y_1'), \ldots, (x_{100}', y_{100}')$.

   For $d$ from 1 to 31, compute the least-squares coefficients of a polynomial of degree $d$ with the same training data as in the last problem using both the QR method and the normal equations.

   For each trained polynomial $p_d(x)$, compute the normalized testing error:
$$e_d = \left( \frac{\sum_{i=1}^{100} (y_i' - p(x_i'))^2}{\sum_{i=1}^{100} (y_i')^2} \right)^{1/2} \tag{2}$$
   Plot the error $e_d$ versus $d$ on a log scale (that is, use `semilogy`). Make sure to include both (i) the error computed using the QR decomposition and (ii) the error computed using the normal equations. Interpret the error behavior. (HINT: It's related to the condition number of the matrix in the least-squares problem.)

BONUS (50 POINTS): Here some US COVID case counts from back in March 2020.

| Days since Feb 29 | Case count |
|---|---:|
| 1 | 89 |
| 2 | 105 |
| 3 | 125 |
| 4 | 159 |
| 5 | 227 |
| 6 | 331 |
| 7 | 444 |
| 8 | 564 |
| 9 | 728 |
| 10 | 1000 |
| 11 | 1267 |
| 12 | 1645 |
| 13 | 2204 |
| 14 | 2826 |
| 15 | 3485 |
| 17 | 7038 |

Derive the linear least-squares system whose solution contains the coefficients of a log-linear model for case count over time. Plot the data on top of the model on a log scale. How well does the log-linear model (which represents exponential growth) appear to model the growth in the case count? What was the case doubling time over this roughly two-week period?